

(19)



JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: 10224897 A

(43) Date of publication of application: 21.08.98

(51) Int. Cl

H04R 25/00

G10L 3/00

G10L 9/00

G10L 9/02

H03G 3/00

(21) Application number: 09026262

(71) Applicant: SONY CORP

(22) Date of filing: 10.02.97

(72) Inventor: MIURA MASAMI

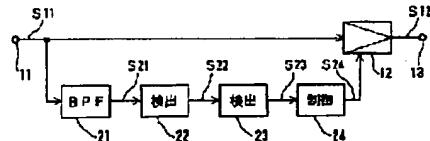
(54) PROCESSING CIRCUIT FOR AUDIO SIGNAL

(57) Abstract:

PROBLEM TO BE SOLVED: To facilitate listening of even continuous sounds in speech while reducing the effect of masking at the time of relay.

SOLUTION: A variable gain amplifier 12 is provided on the signal line of audio signal S11. A detection circuit 23 for detecting the end point of audio signal S11 and a control circuit 24 for controlling the gain of variable gain amplifier 12 according to the detection output of this detection circuit 23 are provided. When the end point of audio signal S11 is detected by the detection circuit 23, at the variable gain amplifier 12, the amplitude of audio signal S11 is enlarged for a prescribed period by the control output of control circuit 24.

COPYRIGHT: (C)1998,JPO



(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平10-224897

(43)公開日 平成10年(1998)8月21日

(51)Int.Cl. ⁶	識別記号	F I	H
H 04 R 25/00		H 04 R 25/00	
G 10 L 3/00	5 1 3	G 10 L 3/00	5 1 3 A
			5 1 3 B
9/00	3 0 1	9/00	3 0 1 A
9/02	3 0 1	9/02	3 0 1 A

審査請求 未請求 請求項の数 7 OL (全 8 頁) 最終頁に統く

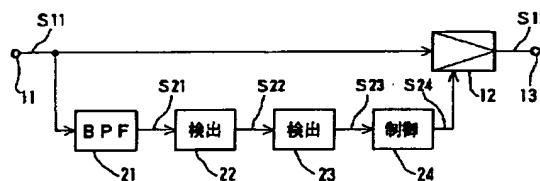
(21)出願番号 特願平9-26262	(71)出願人 ソニー株式会社 東京都品川区北品川6丁目7番35号
(22)出願日 平成9年(1997)2月10日	(72)発明者 三浦 雅美 東京都品川区北品川6丁目7番35号 ソニ ー株式会社内

(54)【発明の名称】 音声信号の処理回路

(57)【要約】

【課題】 錆時マスキングの影響を軽減する。

【解決手段】 音声信号S11の信号ラインに可変利得アンプ12を設ける。音声信号S11の終了点を検出する検出回路23と、この検出回路23の検出出力にしたがって可変利得アンプ12の利得を制御する制御回路24とを設ける。検出回路23により音声信号S11の終了点が検出されたとき、可変利得アンプ12において、制御回路24の制御出力により、音声信号S11の振幅を所定の期間にわたって大きくする。



【特許請求の範囲】

【請求項1】音声信号の供給される可変利得アンプと、上記音声信号の終了点を検出する検出回路と、この検出回路の検出出力にしたがって上記可変利得アンプの利得を制御する制御回路とを有し、上記検出回路により上記音声信号の終了点が検出されたとき、上記可変利得アンプにおいて、上記制御回路の制御出力により、上記音声信号の振幅を所定の期間にわたって大きくするようにした音声信号の処理回路。

【請求項2】請求項1に記載の音声信号の処理回路において、

上記検出回路は、上記音声信号のレベルが音声の存在を認める第1のしきい値を一度越えてから、上記音声の終了と判定する第2のしきい値より小さくなつた時点を、上記検出回路とするようにした音声信号の処理回路。

【請求項3】請求項1あるいは請求項2に記載の音声信号の処理回路において、

上記音声信号をバンドパスフィルタに供給して上記音声のピッチ成分およびフォルマント成分を出力信号として取り出し、

上記検出回路において、上記バンドパスフィルタの出力信号から上記音声信号の終了点を検出するようにした音声信号の処理回路。

【請求項4】請求項1あるいは請求項2に記載の音声信号の処理回路において、

第1のしきい値が第2のしきい値以上であるようにした音声信号の処理回路。

【請求項5】請求項1あるいは請求項2に記載の音声信号の処理回路において、

第2のしきい値を、上記音声が存在する期間における上記音声信号のレベルの最大値と、別に設定した上限値とのうちの小さい方の値の数パーセント～数十パーセントに設定するようにした音声信号の処理回路。

【請求項6】請求項1に記載の音声信号の処理回路において、

上記連続する音声部分のレベルにより上記可変利得アンプの利得を制御するようにした音声信号の処理回路。

【請求項7】請求項1に記載の音声信号の処理回路において、

上記音声信号の終了点からの時間により上記可変利得アンプの利得を制御するようにした音声信号の処理回路。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、補聴器、電話、拡声器、音声通信などの分野で用いられる音声信号の処理回路に関する。

【0002】

【従来の技術】音声を伝送あるいは再生する場合、その伝送系あるいは再生系に残響やエコーが多いと、結果の音声の明瞭度が低下してしまう。そこで、そのようなと

きには、発話速度を遅くする、連続して発声される語音を細かく分解し、時間をあけて再生するなどの処理が行なわれている。

【0003】また、子音のような高域周波数が聞き取りにくいときには、周波数イコライザ処理により高域周波数の強調を行うこともある。さらに、いわゆる維時マスキング（エネルギーの大きい母音と子音とが続くとき、その母音により子音がマスクされる現象）を考慮した重み関数をかける処理も試みられている。

【0004】さらに、以上の処理は難聴者や老人を対象に行われることもある。

【0005】

【発明が解決しようとする課題】ところが、上述したように、発話速度を遅くしたり、連続して発声される語音を分解したりすると、次のような問題点を生じてしまう。

【0006】1. 原音声との間に時間のずれを生じ、即時性がなくなってしまう。したがって、会話などを行なうときには使えない。また、放送などを聞く場合であっても、聞き終わるまでの時間が長くなってしまう。

2. 語音の知覚判断には音声成分の変化速度も重要な手がかりになっているので、発話速度を遅くすると、この手がかりが変化して別な語音に知覚されてしまうことがある。

3. 語音を分解してゆっくり再生すると、語音のまとまりとしての情報や過渡的な変化部分の情報が失われ、明瞭度の悪くなることがある。

4. 常に高域周波数を増幅した音声は、音色のバランスがくずれて不快であったり、聞き取りにくいことがある。

5. 維時マスキングを考慮した重み関数をかける処理は、少なくとも重み関数の時間長の遅延が生じてしまい、即時性が失われてしまう。この結果、口の動きと処理音との間に時間ずれを生じて明瞭度に悪い影響を与えることがある。また、イヤホンからマイクロフォンへの音響的フィードバックがあるときには、その時間遅れによって残響音のような現象が引き起こされてしまう。

【0007】この発明は、以上のような問題点を一掃しようとするものである。

【0008】

【課題を解決するための手段】このため、この発明においては、音声信号の供給される可変利得アンプと、上記音声信号の終了点を検出する検出回路と、この検出回路の検出出力にしたがって上記可変利得アンプの利得を制御する制御回路とを有し、上記検出回路により上記音声信号の終了点が検出されたとき、上記可変利得アンプにおいて、上記制御回路の制御出力により、上記音声信号の振幅を所定の期間にわたって大きくするようにした音声信号の処理回路とするものである。したがって、連続する音声に子音が続くとき、その振幅が拡大される。

【0009】

【発明の実施の形態】ところで、通常の会話の音声は、ある程度のまとまりをもって発音されており、音声の言語知覚も、各音の知覚と、まとまった語音の特徴の知覚との両方から行なわれていると言われている。

【0010】また、健聴者の場合、雑音が少なく受聴環境の良いところでは、特別な音声処理を行なわなくても音声を十分に聞き取ることができるが、雑音が多いところでは、言葉の聞き取りが悪くなったりする。この原因はいくつか考えられるが、主なものに、継時マスキングの影響がある。つまり、前の母音が次の音の子音部をマスクすることができ、この結果、子音の聴覚的な感度が悪くなってしまい聞き取りにくくなるものである。

【0011】そこで、この発明においては、音声のまとまりの終了点を検出し、この終了点から十数m秒～数十m秒の期間、音声信号（特にその高域成分）の振幅を増幅し、これにより子音に対する聴覚的な感度を相対的に高くするものである。また、音声のまとまりの終了点は、音声のピッチ成分およびフォルマント成分のレベル変化を解析することにより検出する。

【0012】図1は、この発明の一形態を示すもので、もとの処理前の音声信号S11が入力端子11を通じて可変利得アンプ12に供給され、このアンプ12の出力信号S12が出力端子13に取り出される。

【0013】さらに、端子11の信号S11が、前処理のため、バンドパスフィルタ21およびレベル検出回路22に順に供給される。この場合、バンドパスフィルタ21は、連続する音声の終了点を検出しやすくし、かつ、雑音による影響が小さくなるように、信号S11からピッチ成分とフォルマント成分とを抽出するものである。したがって、その通過帯域は、例えば150Hz～1000Hzとされている。

【0014】また、レベル検出回路22は、バンドパスフィルタ21の出力信号S21を使用して連続する音声の終了点を検出するためのものである。このため、レベル検出回路22は、例えば、信号S21を両波整流するとともに、その低域成分（例えば60Hz以下の成分）を取り出すことにより、信号S11のレベルを示す信号S22を形成している。

【0015】そして、このレベル検出回路22の検出信号S22が終了点検出回路23に供給され、連続する音声の終了点が検出され、その終了点の検出信号S23が利得制御回路24に供給されて制御信号S24が形成され、この信号S24が可変利得アンプ12に利得の制御信号として供給される。

【0016】この場合、連続する音声の終了点の検出は、音声レベル、つまり、信号S22のレベルが、第1のしきい値（音声存在判定しきい値）を一度越え、その後、第2のしきい値（音声終了判定しきい値）よりも小さくなるときに行う。また、利得の制御は、その音声の

終了点の検出時点の直後から、例えば十数m秒～数十m秒の間にわたって、利得が大きくなるように行う。ただし、第1のしきい値は第2のしきい値以上とする。

【0017】このような構成によれば、端子11に入力された音声信号S11が連続している期間は、制御信号S23により可変利得アンプ12の利得は基準値に固定されている。したがって、その音声信号S11が、そのまま出力信号S12として端子13に取り出される。

【0018】しかし、音声信号S11の連続が終了すると、その終了点の時点から制御信号S24により可変利得アンプ12の利得が基準値よりも大きくなるので、その終了点の時点から所定の期間、音声信号S11があれば、音声信号S12の振幅は本来の大きさよりも大きくなる。したがって、連続する音声（音声信号S11）の直後の子音に対する聴覚的な感度が継時マスキングにより低下していく中、その子音の振幅は本来の大きさよりも大きくなっているので、その感度の低下は相殺されることになり、その子音を含む音声の明瞭度が向上する。

【0019】図2～図4は、検出回路23および制御回路24が、検出信号S22から制御信号S24を形成する方法の一形態を示す。すなわち、この場合には、図1に示した回路の全部がデジタル化され、例えばDSPにより構成される。そして、音声信号S11はもとの処理前のアナログ音声信号をA/D変換したデジタル音声信号とされる。

【0020】そして、検出回路23および制御回路24においては、デジタル音声信号S11の1サンプルごとに、図2の処理ルーチン100が実行される。このルーチン100においては、アンプ12の利得を変更するとき、音声信号S12のレベルが不連続にならないようにするため、例えば図5に示すように、次第に変化するようになっている。

【0021】また、ルーチン100および以下の説明において、各変数の意味は以下のとおりである。

【0022】

e(i) : 音声信号S11の第*i*番目のサンプルのレベル。

peak : 音声信号S11のピーク値。

peakmax : 値peakの上限を制限するときの値。

rate : 信号S11が値peakから何%小さくなったら、音声がなくなったと判断するかの割り合い。

voicemin : 第1のしきい値（音声存在判定しきい値）。信号S11がこの値を越えたとき、音声の存在を認める。

threshold : 第2のしきい値（音声終了判定しきい値）。信号S11がこの値よりも小さくなったらとき、音声の終了点であると判定する。voicemin≥threshold、threshold=peak×rateである。

slope1 : アンプ12の利得を大きくしていく期間T1の長さ。

slope2 : アンプ1 2の利得をもとに戻していく期間T2 の長さ。

enable : 第(i-1) 番目のサンプルに、音声が存在するかどうかを示すフラグ。

enable= "0" のとき、音声は存在しない。

enable= "1" のとき、音声が存在する。

jj : 連続する音声の終了点から時間を計数するときの変数。

【0023】そして、ルーチン100においては、まず、ステップ101において、 $e(i) < threshold$ であるかどうかをチェックすることにより第i 番目のサンプルに音声が存在しないかどうかが判別され、信号S11が存在しないとき ($e(i) < threshold$ のとき) には、処理はステップ102に進む。そして、ステップ102において、enable= "1" であるかどうかをチェックすることにより1つ前のサンプルである第(i-1) 番目のサンプルに音声が存在していたかどうかが判別され、1つ前のサンプルに音声が存在していたとき (enable= "1" のとき) には、処理はステップ103に進む。

【0024】したがって、処理がステップ103に進むのは、連続する2つのサンプルのうち、前のサンプルには音声が存在し、かつ、後ろのサンプルには音声が存在しないときであり、これは連続した音声が終了したときである。つまり、連続した音声の終了が検出されたときである。

【0025】そこで、ステップ103において、次の第(i+1) 番目のサンプルに備えてenable= "0" とされるとともに、jj=0 とされ、第i 番目のサンプルについての処理を終了する。

【0026】また、ステップ102において、1つ前のサンプルに信号S11が存在していないとき (enable= "0" のとき) には、処理はステップ111に進み、 $jj < slope1$ であるかどうかをチェックすることによりサンプル時点が期間T1 に含まれるかどうかが判別され、含まれるとき ($jj < slope1$ のとき) には、処理はステップ112に進む。つまり、期間T1 には、処理はステップ112に進む。

【0027】そして、このステップ112において、アンプ1 2の利得が1ステップ分だけ大きくなる。また、変数jjが「1」だけインクリメントされる。そして、これで第i 番目のサンプルについての処理を終了する。

【0028】さらに、ステップ111において、サンプル時点が期間T1 に含まれないとき ($jj \geq slope1$ のとき) には、処理はステップ121に進む。つまり、期間T1 を過ぎているときには、処理はステップ121に進む。

【0029】そして、このステップ121において、 $jj \geq slope1$ 、かつ、 $jj < (slope1 + slope2)$ であるかどうかをチェックすることによりサンプル時点が期間T2 に

含まれるかどうかが判別され、含まれるとき ($jj \geq slope1$ 、かつ、 $jj < (slope1 + slope2)$ のとき) には、処理はステップ122に進む。つまり、期間T2 には、処理はステップ122に進む。

【0030】そして、このステップ122において、アンプ1 2の利得が1ステップ分だけ小さくなる。また、変数jjが「1」だけインクリメントされる。そして、これで第i 番目のサンプルについての処理を終了する。

【0031】さらに、ステップ121において、サンプル時点が期間T2 を過ぎているとき ($jj \geq (slope1 + slope2)$ のとき) には、アンプ1 2の利得の変更などをしないで、第i 番目のサンプルについての処理を終了する。

【0032】こうして、以上の処理によれば、連続していた音声が終了すると、これが検出され、図3に示すように、アンプ1 2の利得が制御される。

【0033】そして、ルーチン100においては、さらに、アンプ1 2の利得を大きくしたときに、音声レベルのピーク値が極端に大きくなることを防ぐため、ピーク値に上限を設けている。また、ルーチン100においては、第2のしきい値は、音声の連続ごとに設定するとともに、音声レベルが第1のしきい値を越えてから音声レベルのピークを見つけ、そのピーク値の数%～数十%の値に設定している。

【0034】すなわち、ステップ101において、音声が存在するとき ($e(i) \geq threshold$ のとき) には、処理はステップ131に進む。そして、このステップ131において、enable= "0" であるかどうかをチェックすることにより1つ前のサンプルである第(i-1) 番目のサンプルに音声が存在していないかどうかが判別され、1つ前のサンプルに音声が存在していないとき (enable= "0" のとき) には、処理はステップ141に進む。

【0035】この場合、処理がステップ141に進むのは、ステップ101およびステップ131を通じてであるから、これは、連続する2つのサンプルのうち、前のサンプルに音声が存在しなくて後ろのサンプルに音声が存在する場合であり、つまり、音声が開始されたときである。

【0036】そこで、ステップ141において、 $e(i) > voicemin$ であるかどうかをチェックすることにより音声が開始されたかどうかの確認が取られ、開始されているとき ($e(i) > voicemin$ のとき) には、処理はステップ142に進み、レベル $e(i)$ が上限値 $peakmax$ を越えているかどうかが判別される。そして、越えていないとき ($e(i) \leq peakmax$ のとき) には、処理はステップ143に進み、ピーク値 $peak$ がレベル $e(i)$ に設定され、その後、処理はステップ145に進む。

【0037】また、ステップ142において、レベル $e(i)$ が上限値 $peakmax$ を越えているとき ($e(i) > peakmax$

のとき)には、処理はステップ144に進み、ピーク値peakが上限値peakmaxに設定され、その後、処理はステップは145に進む。

【0038】そして、ステップ145においては、ステップ143あるいはステップ144で設定されたピーク値peakにしたがって、第2のしきい値thresholdが、threshold = peak × rateで示される値に設定されるとともに、次の第(i+1)番目のサンプルに備えてenable = "1"とされ、その後、ステップ161に進む。

【0039】さらに、ステップ141において、音声が開始されていないとき($e(i) \leq voicemin$ のとき)には、処理はステップ141からそのままステップ161に進み、ピーク値peakおよびフラグenableは設定されない。

【0040】こうして、音声が開始されたときには、その開始時のレベルに対応して第2のしきい値thresholdが設定される。

【0041】一方、ステップ131において、1つ前のサンプルに音声が存在していたとき(enable = "1"のとき)には、処理はステップ151に進む。この場合、処理がステップ151に進むのは、ステップ101およびステップ131を通じてであるから、これは、連続する2つのサンプルの両方に音声が存在するときであり、音声が連続しているときである。つまり、音声が連続している期間のときである。

【0042】そこで、ステップ151において、第i番目のサンプルのレベルe(i)がそれまでのピーク値peakと比較され、レベルe(i)のほうが大きいとき($e(i) > peak$ のとき)には、処理はステップ151からステップ152に進み、レベルe(i)が上限値peakmaxを越えているかどうかが判別される。そして、越えていないとき($e(i) < peakmax$ のとき)には、処理はステップ153に進み、ピーク値peakがレベルe(i)に更新され、その後、処理はステップ155に進む。

【0043】また、ステップ152において、レベルe(i)が上限値peakmaxを越えているとき($e(i) < peakmax$ のとき)には、処理はステップ154に進み、ピーク値peakが上限値peakmaxに更新され、その後、処理はステップ155に進む。

【0044】そして、ステップ155においては、ステップ153あるいはステップ154で更新されたピーク値peakにしたがって、第2のしきい値thresholdが、threshold = peak × rateで示される値に更新され、その後、ステップ161に進む。

【0045】さらに、ステップ151において、それまでのピーク値peakのほうが大きいとき($e(i) \leq peak$ のとき)には、処理はステップ151からそのままステップ161に進み、ピーク値peakは更新されない。

【0046】こうして、音声が連続しているときには、その連続期間におけるピーク値peakが見つけられるとともに、そのピーク値peakにしたがって第2のしきい値th-

resholdが更新される。

【0047】そして、処理がステップ161に進むと、 $jj > 0$ かつ、 $jj < slope1$ であるかどうかをチェックすることによりサンプル時点が期間T1に含まれるかどうかが判別され、含まれるとき($jj > 0$ かつ、 $jj < slope1$ のとき)には、処理はステップ162に進む。つまり、期間T1には、処理はステップ162に進む。

【0048】そして、このステップ162において、アンプ12の利得が1ステップ分だけ大きくされる。また、変数jjが「1」だけインクリメントされる。そして、これで第i番目のサンプルについての処理を終了する。

【0049】さらに、ステップ161において、サンプル時点が期間T1に含まれないとき($jj \geq slope1$ のとき)には、処理はステップ171に進む。つまり、期間T1を過ぎたときには、処理はステップ171に進む。

【0050】そして、このステップ171において、 $jj \geq slope1$ かつ、 $jj < (slope1 + slope2)$ であるかどうかをチェックすることによりサンプル時点が期間T2に含まれるかどうかが判別され、含まれるとき($jj \geq slope1$ かつ、 $jj < (slope1 + slope2)$ のとき)には、処理はステップ172に進む。つまり、期間T2には、処理はステップ172に進む。

【0051】そして、このステップ172において、アンプ12の利得が1ステップ分だけ小さくされる。また、変数jjが「1」だけインクリメントされる。そして、これで第i番目のサンプルについての処理を終了する。

【0052】さらに、ステップ171において、サンプル時点が期間T2を過ぎているとき($jj \geq (slope1 + slope2)$ のとき)には、処理はステップ181に進み、変数jjが「0」に初期化され、これで第i番目のサンプルについての処理を終了する。

【0053】以上の処理により、連続する音声の終了点を起点として十数m秒～数十m秒の期間、音声信号S11の振幅が大きくなるので、その連続する音声(音声信号S11)の直後の子音に対する聽感的な感度が維持マスキングにより低下していても、その感度の低下は、信号S11の振幅の増大により相殺されることになり、その子音を含む音声の明瞭度が向上する。

【0054】図4は、音声波形の観測結果を示すもので、図4Aはルーチン100による処理を行っていない音声信号S11の波形、図4Bはルーチン100による処理を行った音声信号S12の波形である。また、このときの発声内容は、「1行目に書いてください」である。

【0055】そして、矢印A、E、Fにより示すように、連続する音声の終了点から次の子音までの時間間隔が短いときには、その子音に対する維持マスキングが大きいが、その子音の振幅は拡大されて強調されている。また、矢印B、C、Dにより示すように、次の子音まで

の時間間隔が長いときには、維時マスキングは小さいが、これに対応して子音の振幅は強調されていない。

【0056】したがって、上述の処理回路によれば、音声を残響やエコーなどのある系で伝送あるいは再生するとき、あるいは難聴者や老人が音声を聞くとき、以下のような効果を得ることができる。

1. 次に発声される子音への維時マスキングが軽減されるように、その子音だけを強調しているので、音声がはっきりし、明瞭度を改善できる。

2. 常に音声の高域を強調すると、音色のバランスが崩れたような不快感を伴うが、維時マスキングが起きているときのみ子音を強調するので、そのような不快感がない。

3. 原理的に即時処理ができるので、発声者の口の動きと処理音との間に時間差の生じることがない。また、イヤホンからマイクロフォンへの音響的フィードバックがあっても、残響音のような音にはならないので、聞きやすい。

4. 語音の知覚判断にとって重要な音声成分の変化速度や、語音のまとまりとしての情報および過渡的な変化部分の情報が失われない。

【0057】なお、上述において、維時マスキングによるマスキング量は、妨害音として作用する音声部分（連続する音声区間）のレベルおよび維続時間と、連続する音声部分の終了点からの時間によって変化するが、もとの音声信号S11の振幅が制御される期間およびその大きさを、維時マスキングのマスキング量にあわせて加減す

ることもでき、マスキング量が大きいときには、振幅およびその制御期間を大きくすればよい。

【0058】また、上述においては、可変利得アンプ12において、音声信号S11の全帯域について振幅を大きくしているが、子音に対応する高域だけ振幅を大きくしてもよい。

【0059】

【発明の効果】この発明によれば、音声がはっきりし、明瞭度を改善できる。また、常に音声の高域を強調するときのような不快感がない。さらに、発声者の口の動きと処理音との間に時間差の生じることがない。

【0060】また、イヤホンからマイクロフォンへの音響的フィードバックがあっても、残響音のようにならず、聞きやすい。さらに、語音の知覚判断などに有効な情報が損なわれない。

【図面の簡単な説明】

【図1】この発明の一形態を示す系統図である。

【図2】この発明の一形態の一部を示すフローチャートである。

【図3】図2の続きを示すフローチャートである。

【図4】図3の続きを示すフローチャートである。

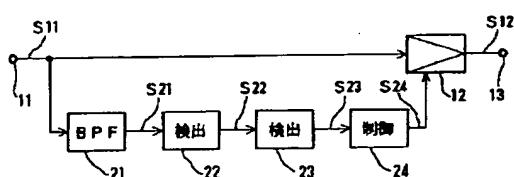
【図5】この発明を説明するための図である。

【図6】この発明を説明するための図である。

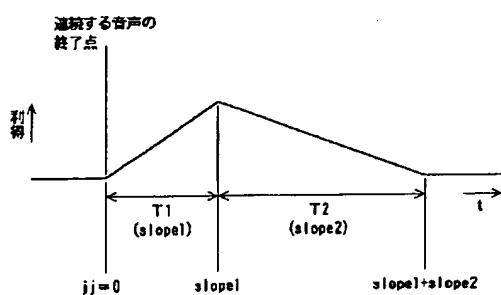
【符号の説明】

12…可変利得アンプ、21…バンドパスフィルタ、22…レベル検出回路、23…終了点検出回路、24…利得制御回路、100…処理ルーチン

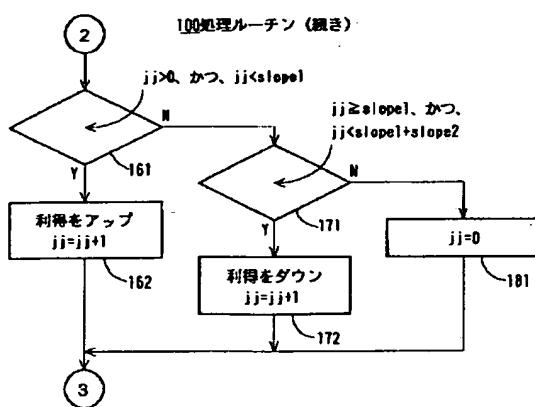
【図1】



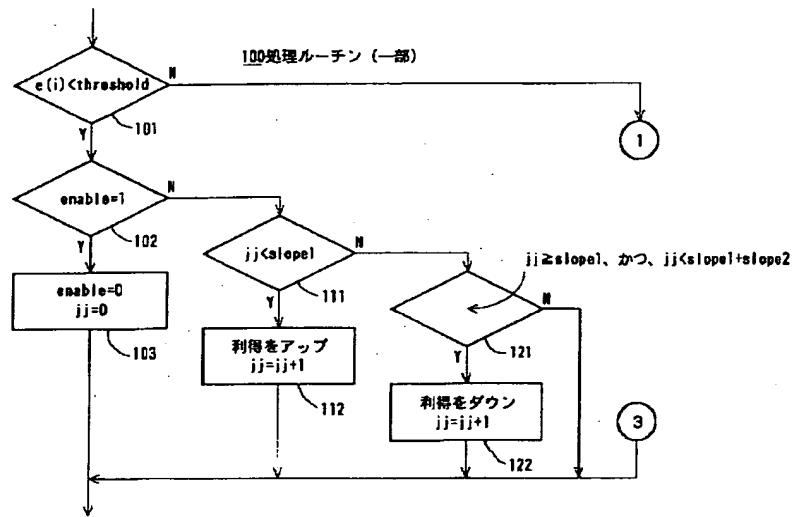
【図5】



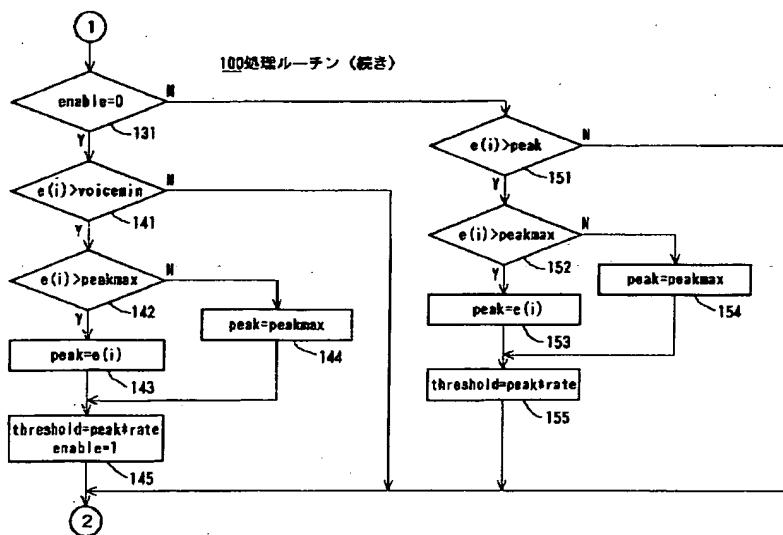
【図4】



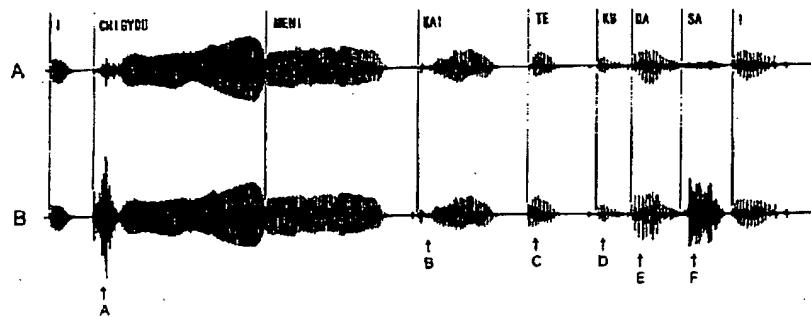
【図2】



【図3】



【図6】



フロントページの続き

(51) Int. Cl.⁶
H 03 G 3/00

識別記号

F I
H 03 G 3/00

A